The information below will be used for the entire Statistics exam. Add data are on a tear-off page in the back of the exam for reference.

Ms. Raskin found the GPA quality points for the first quarter for her Alpha Statistics class.
(Note: A = 4pts, B = 3pts, etc.) They are as follows:

0, 0, 0, 1.7, 2, 2, 2, 2, 2.3, 2.3, 2.3, 2.3, 2.7, 2.7, 2.7, 2.7, 3, 3, 3.3, 3.3, 4

Q1 Med Q3

1) Find the **median** and **IQR** of these data. (2pts)

median = 2.3

IQR = 2.85 - 2 = 0.85

2) Find the **mean** of these data. (2pts)

Sum = 46.3

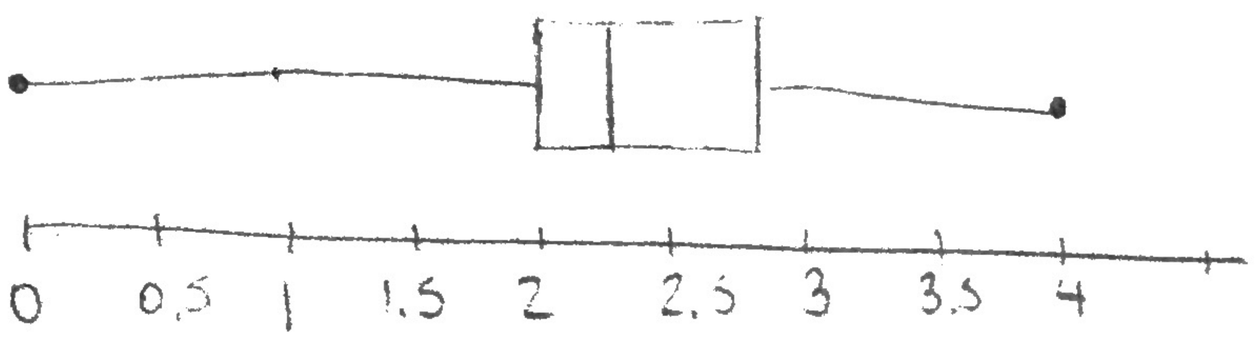
46.3 ÷ 21 = 2.205

3) The **standard deviation** of these data is 1.0703. What does this tell you about spread of the data? (2pts)

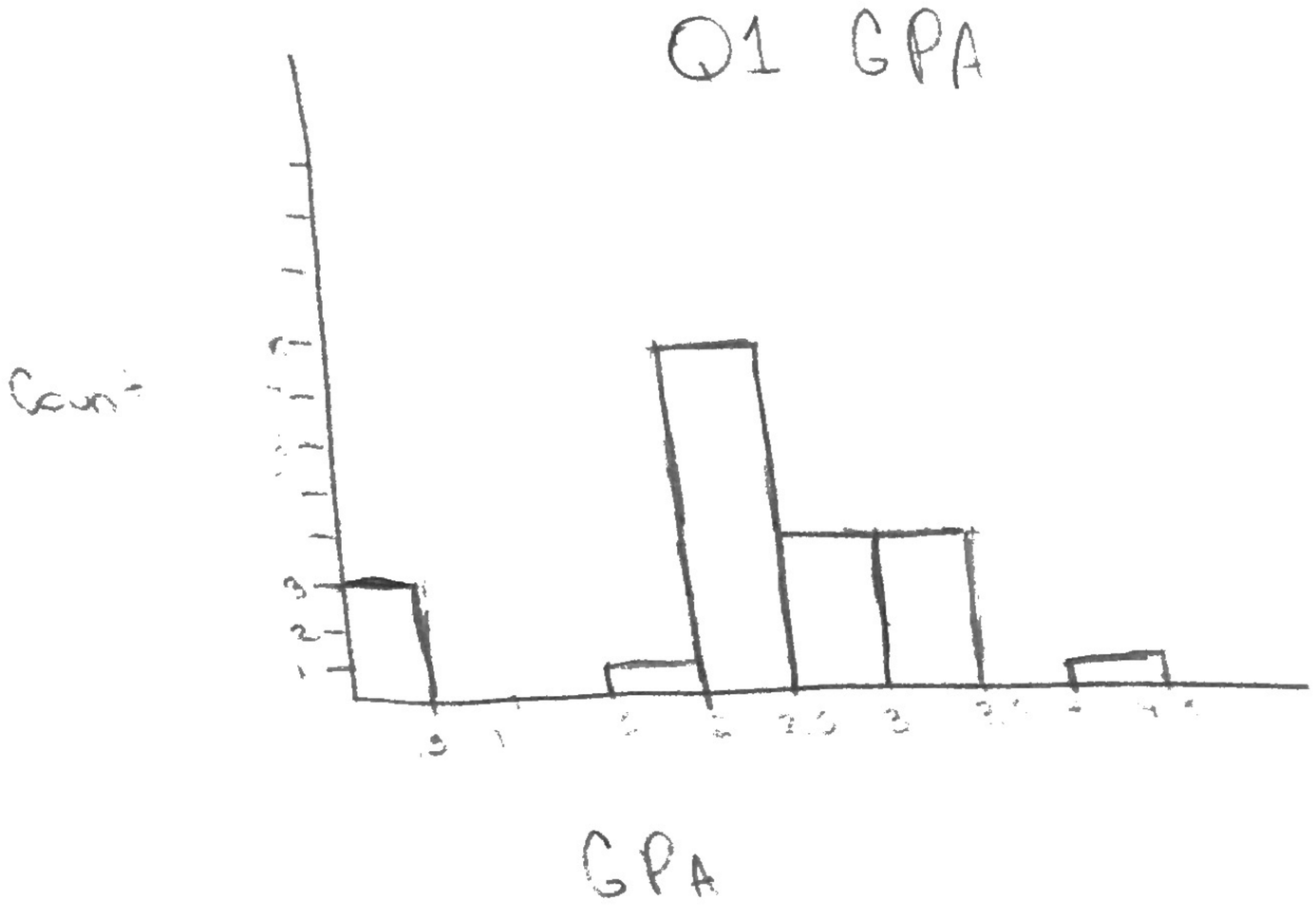
The average distance from the mean GPA is 1.0703 points.

4) Make a **boxplot** of these data. (4pts)

Q1 GPA



5) Make a **histogram** of these data. (4pts)



6) Describe your histogram using SOCS (**s**hape, **o**utliers, **c**enter, and **s**pread). (2pts)

The shape is relatively symmetric, though there is a gap between 0.5 and 1.5. The zeroes are outliers. The median is 2.3 and there is a good-sized spread of approximately 1 point.

7) Are any of these grades **outliers**? If so, explain why they are. If not, explain why none are outliers. (2pts)

The zeroes are outliers because they are more than 1.5 IQRs below Q1.

8) Are earning a 2.3 GPA point grade and a 3.3 GPA point grade **mutually exclusive/disjoint**? Why or why not? (1pt)

They are mutually exclusive. You cannot have both GPAs at the same time.

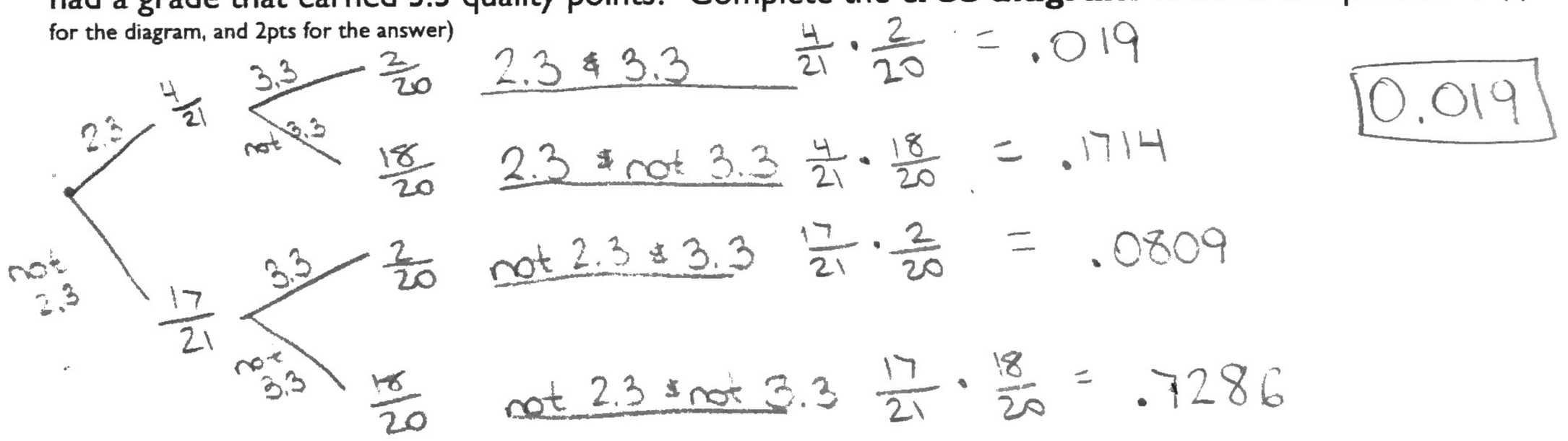
9) Ms. Raskin randomly selects one student from the class to be in charge while she finds extra calculators for the seniors who didn't bring theirs. Which student is **more likely** to be chosen: a student whose grade earned 2.3 GPA quality point or a student whose grade earned 3.3 quality points? **Why?** (2pts)

The student with a 2.3 GPA is more likely to be chosen because there are 4 of them and only 2 students with a 3.3 GPA.

10) **What is the probability** that the first student she chooses was a student whose grade earned 2.3 quality GPA point? (2pts)

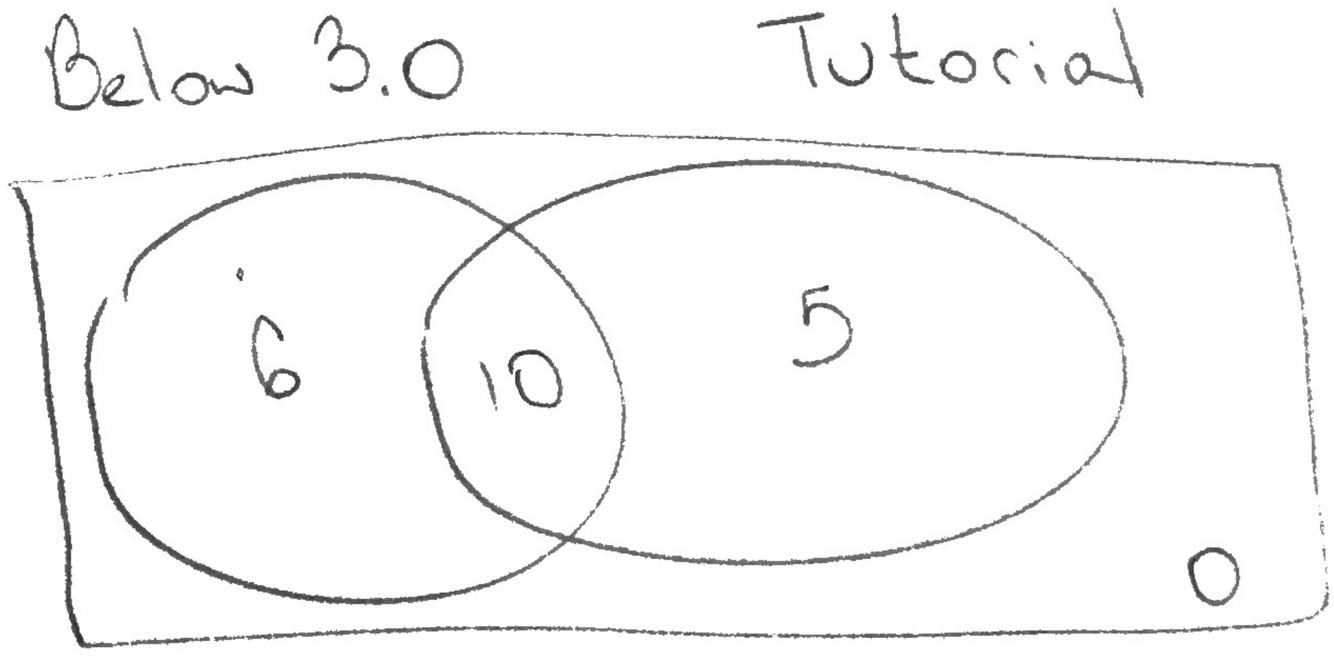
$\frac{4}{21} = .19$

11) There are a lot of students in the class, so Ms. Raskin selects a second student to help be in charge. What is the probability that the first student had a grade of 2.3 quality point **AND** the second student had a grade that earned 3.3 quality points? Complete the **tree diagram** to solve this problem. (3pts for the diagram, and 2pts for the answer)



0.019

12) 16 students earned below 3 GPA quality points. 15 students attended tutorial. Only 10 students did both. How many students earned below 3 GPA quality points and did not attend tutorial? Draw a **Venn diagram** to solve this problem. (3pts for the diagram and 2pts for the answer)



6 students earned below 3.0 and did not attend tutorial.

Statistics Final Exam

The parallel boxplots below show the Q1 grades for school years 2011-2012 through 2015-2016. (This is also on the tear-out sheet on the back.)

13) Choose two measures of SOCS on which to compare **all** of the ~~five~~ four boxplots. (4pts)

14) Which class had an **outlier**? What does this mean? (2pt)

2014-2015 had an outlier. That means that one person had a much lower grade (F) as compared to the rest of the class.

15) How can a grade of 4 GPA points (a grade of A) be an **outlier** in one class and a **maximum** in another class? (This did not happen in this case but is a hypothetical event.) (3pts)

In one class, only one person might have an a, while the other people have very low grades. That makes that person an outlier. In another class, the highest grade might be an A, but several people might get that grade.

16) Compare the **medians** of the distributions. Make sure to tell what this means about the classes. (2pts)

The medians of the classes of 2011-2012, 2012-2013, and 2013-2014 are pretty much the same. The class of 2014-2015 had a much lower median. In general, that class's grades were lower.

17) Are any distributions **skewed**? Which direction (toward the lower grades or toward the higher grades)? What does this tell you about the grades in those classes? (2pts)

The 2014-2015 class was significantly right skewed (toward higher grades). Most of these students earned low grades, but a few earned higher grades. 25% of the class earned between a 3 and a 4, whereas almost 50% of students in other years earned these grades.

18) **Compare** the **IQR** of 2014-2015 with the **IQR** of 2012-2013. Make sure to tell what this means. (2pts)

The IQR of 2014-2015 was significantly smaller than that of 2012-2013. Half of the students earned grades between 2.2 & 2.7.

19) On average, which class had the highest grades? Explain why you chose this class. (2pts)

On average, 2014-2015 had the highest grades. 75% of those students earned between a 2.2 (C+) and 4.0 (A).

20) Ms. Raskin thought that the students' **Q1 grade** could predict their **final exam grade**. Identify the **explanatory variable** in this relationship. (1pt)

Q1 grade

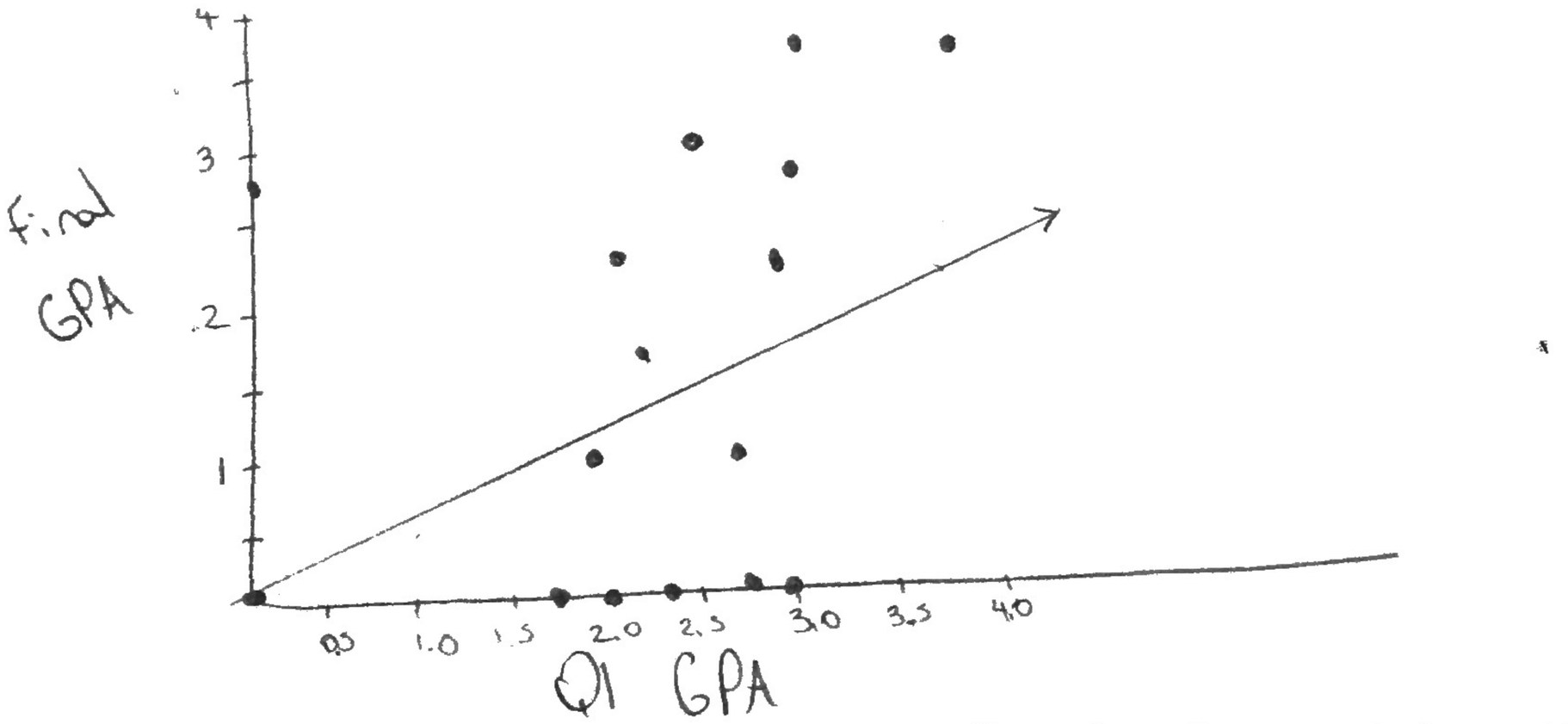
21) Identify the **response variable** in this relationship. (1pt)

final exam grade.

The table below shows the Q1 GPA points versus the Y1 GPA points by student.

Q1 GPA Points	Final Exam GPA Points
2.3	0
2.7	1
0	0
2	1
2.7	0
2.7	3
4	3.7
2	0
0	0
2.3	2.3
2	0
3	2.3
0	2.7
3.3	2.7
2.7	0
3	0
2.3	0
2	0
2.3	1.7
1.7	0
3.3	3.7

22) Make a **scatterplot** of these data. (4pts)



23) Describe your data **in context** using **direction, form**, and **strength**. (3pts)

There is a modesate, positive, linear relationship between Q1 GPA and final GPA.

24) The **correlation** of these data is r = 0.3921. What does this tell you about the **relationship** between Q1 and final exam grades? (2pts)

The strength of the relationship is not that strong.

25) The **least square regression line (LSRL)** for these data is $\hat{y} = 0.023 + 0.51x$. Another way to say this is that $\widehat{exam\ grade} = 0.023 + 0.51(Q1\ grade)$. **Why is there a "^" over the "y"** (or over the "exam grade")? (2pts)

The ^ shows that this is a predicted, not actual, y-value.

26) **Draw** this line on your scatterplot. (1pts)

27) Interpret the **slope** in **context**. What does this number mean for this line? (3pts)

For each increase of 1 point in Q1 GPA, we predict a 0.51 increase in final exam GPA.

28) Interpret the **y-intercept** in **context**. What does this number mean for this line? (3pts)

When Q1 GPA is zero, we predict that the final GPA will be 0.023.

29) One student who earned 0 GPA quality points in Q1 worked very hard and earned 2.7 GPA quality points for the exam. What is this student's **residual**? (2pts)

$\hat{y} = 0.023 + 0.51(0)$

$\hat{y} = 0.023$

2.7 ~~Q6~~ 6 9

-0.023

2.697 is the residual

30) What can you say about the Y1 grade of a student whose **residual** is ***negative***? (2pts)

If a student's residual is negative, he or she scored below expected for the Y1 grade.

31) One student notices that her grade had a **negative residual** but a **positive z-score**.

a. What does a **negative residual** tell you about her score? (2pt)

She scored below what was predicted based on her Q1 grade.

b. What does a **positive z-score** tell you about her score? (2pt)

A positive Z-score means that she scored above average.

c. If you could have only one of the following in this situation: a positive residual or a positive z-score, which would you choose? Make sure that you can explain your choice. (3pts)

Answers may vary.

A positive z-score means that a score is above average.

A positive residual means that a score is higher than predicted based on a Q1 grade.

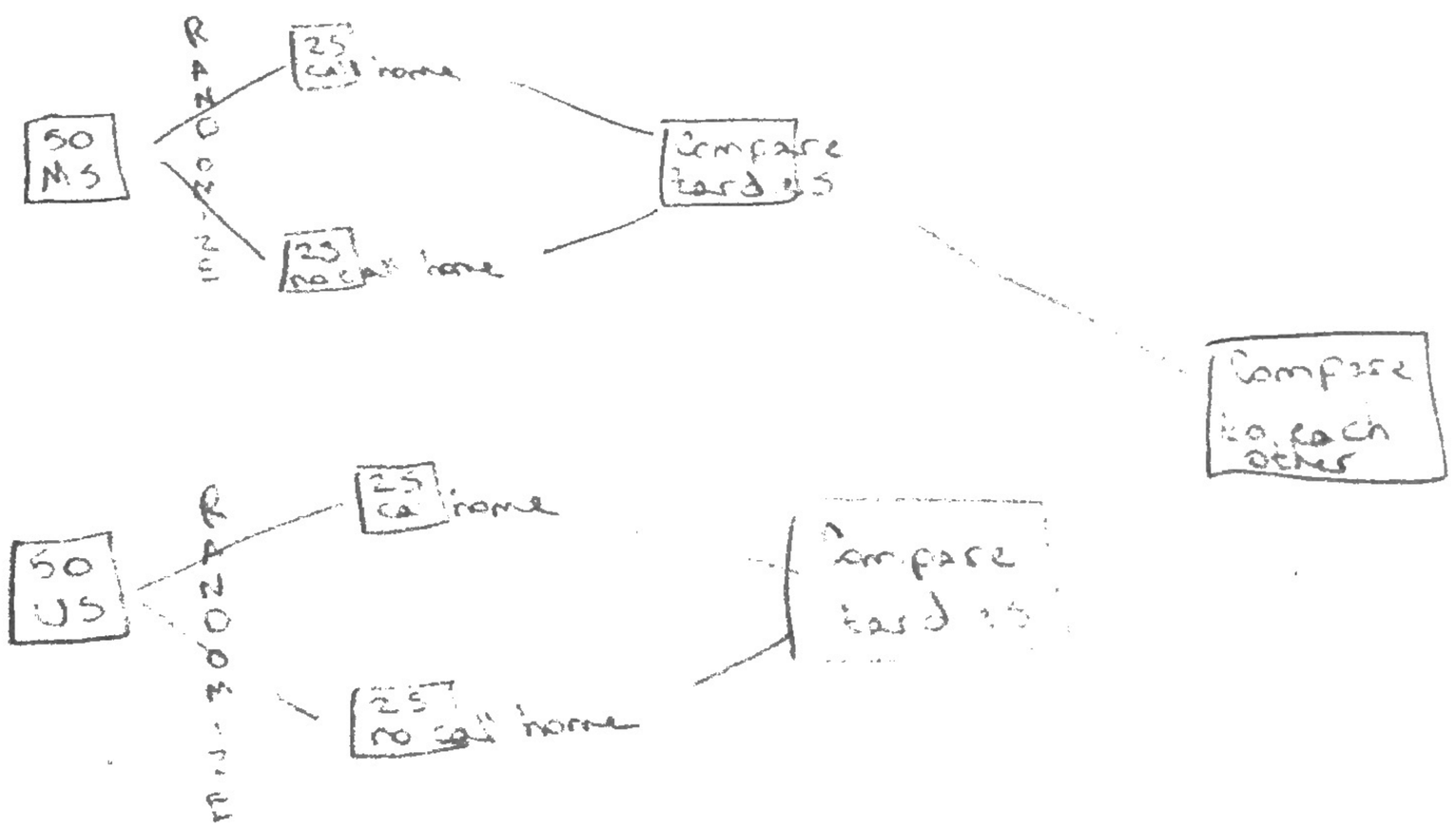
32) Ms. Smith asks Ms. Raskin to look at the students' grades in relation to their English grades. She wonders why the students who earn good grades in English also earn good grades in Statistics. She believes that earning good grades in English **causes** students to earn good grades in Statistics. Can she determine this? Why or why not? (2pt)

Correlation does not equal causation!

33) Ms. Smith asks Ms. Raskin to perform an experiment to see if calling home for class tardies reduces the number of tardies **in the entire school** (middle and upper). Which would be the optimal design/type of experiment to run: **randomized controlled, blocked,** or **matched pairs**? (2pt)

We should design a blocked experiment with blocking based on age (middle vs. upper school) because there will likely be a difference between how middle & upper schoolers' responses to phone calls home.

34) Explain how you would set up an experiment to determine if calling home will reduce the number of class tardies. You may draw a diagram to help, if you want. (4pts)



35) What are the **three principles of experimental design** that Ms. Raskin must follow? (3pts)

I. Control
II. Randomization
III. Replication